

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-66691

(P2000-66691A)

(43)公開日 平成12年3月3日(2000.3.3)

(51)Int.Cl. ⁷	識別記号	FI	テマコード [*] (参考)
G10L	11/00	G10L 7/02	A
	11/06	9/00	C
	19/00	9/16	
		9/18	H

審査請求 未請求 請求項の数12 OL (全 15 頁)

(21)出願番号 特願平10-235543

(22)出願日 平成10年8月21日(1998.8.21)

(71)出願人 000001214

ケイディディ株式会社

東京都新宿区西新宿2丁目3番2号

(72)発明者 中島 康之

東京都新宿区西新宿2丁目3番2号 国際
電信電話株式会社内

(72)発明者 菅野 勝

東京都新宿区西新宿2丁目3番2号 国際
電信電話株式会社内

(74)代理人 100084870

弁理士 田中 香樹 (外1名)

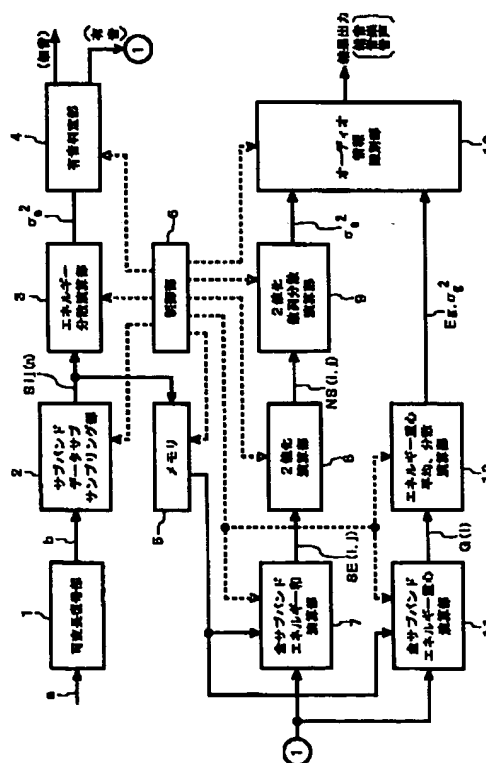
最終頁に続く

(54)【発明の名称】 オーディオ情報分類装置

(57)【要約】

【課題】 簡単かつ高速に、無音／有音区間の判別、音楽区間と音声区間、あるいは音楽区間と音声区間と雑音区間に分類することを可能とするオーディオ情報分類装置を提供することにある。

【解決手段】 有音判定部4はエネルギー分散演算部3で求められた値 σ_e^2 が閾値より大きい時、有音と判定する。有音と判定されると、メモリ5に格納されていたオーディオ情報が読み出されて、全サブバンドエネルギー和演算部7とサブバンドエネルギー重心演算部11に入力される。前記演算部7の出力は2値化演算部8で2値化され、2値化数列分散演算部9で2値化数列の単位時間内の分散 σ_s^2 が求められる。一方、エネルギー重心平均・分散演算部12はエネルギー重心平均 Eg と分散 σ_g^2 を求める。オーディオ情報識別部10は、前記分散 σ_s^2 、エネルギー重心平均 Eg 、および分散 σ_g^2 に対して識別関数を用いて、雑音、音楽、音声の判別を行う。



【特許請求の範囲】

【請求項1】 オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、
入力されたオーディオ情報から単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、
抽出した単位時間ごとの周波数データのエネルギーの分散を求め、分散値の大きさにより無音／有音区間を判定する無音／有音判定手段とを具備することを特徴とするオーディオ情報分類装置。

【請求項2】 請求項1に記載のオーディオ情報分類装置において、

前記オーディオ周波数データ抽出手段によって抽出される単位時間ごとの周波数データは、入力されたオーディオ情報がMPEGデータである場合、単位時間分のMPEG符号化データにおける最低周波数成分のエネルギーの分散を利用することを特徴とするオーディオ情報分類装置。

【請求項3】 オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、
入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と有音区間における音の疎密度により音声であるか音楽であるかを判定する音声／音楽区間判定手段とを具備することを特徴とするオーディオ情報分類装置。

【請求項4】 請求項3に記載のオーディオ情報分類装置において、
疎密度はオーディオ信号のエネルギーの大きさによって2値化された数列の分散を用いて疎密度を判定することを特徴とするオーディオ情報分類装置。

【請求項5】 請求項4に記載のオーディオ情報分類装置において、

前記オーディオ信号のエネルギーは、入力されたオーディオ情報がMPEGデータである場合、単位時間分のMPEG符号化データにおける全周波数成分のエネルギー和を利用することを特徴とするオーディオ情報分類装置。

【請求項6】 請求項3ないし請求項5のいずれかに記載のオーディオ情報分類装置において、

前記音声／音楽区間判定手段は疎密度を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音楽と音声区間の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて各音楽区間と音声区間の判別を行うことを特徴とするオーディオ情報分類装置。

【請求項7】 オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、
入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と入力されたオーディオ情報から有音時の単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、
オーディオ周波数データから単位時間における周波数の重心の平均と重心の標準偏差を求め、周波数の重心の分

布により雑音区間か否かを判別する雑音区間抽出手段を具備することを特徴とするオーディオ情報分類装置。

【請求項8】 請求項7に記載のオーディオ情報分類装置において、

前記オーディオ周波数データ抽出手段によって抽出される単位時間ごとの周波数データは、入力されたオーディオ情報がMPEGデータである場合、単位時間分のMPEG符号化データにおける周波数成分のエネルギーの重心を利用することを特徴とするオーディオ情報分類装置。

10 【請求項9】 請求項7又は8に記載のオーディオ情報分類装置において、

前記雑音抽出手段は、周波数成分の重心の平均と分散を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して雑音と雑音以外の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて各雑音区間と非雑音区間の判別を行うことを特徴とするオーディオ情報分類装置。

【請求項10】 オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、

20 入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と、

入力されたオーディオ情報から有音時の単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、

オーディオ周波数データの単位時間における疎密度および単位時間における周波数の重心の平均と重心の標準偏差を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音声と音楽と雑音の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて音声、音楽、雑音区間の判別を行うことを特徴とする音声／音楽／雑音区間判別手段を具備することを特徴とするオーディオ情報分類装置。

30 【請求項11】 オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、

入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と、

入力されたオーディオ情報から有音時の単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、

40 オーディオ周波数データの単位時間における周波数の重心の平均と重心の標準偏差を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して雑音と雑音以外の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて雑音と雑音以外の区間の判別を行うことを特徴とする雑音区間判別手段と、

オーディオ周波数データの単位時間における疎密度を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音声と音楽の共分散行列を求めておき、前記、
50 雑音区間判別手段で雑音以外と判別された区間に対し

て、正規分布パターンにおけるBayes 決定識別関数を用いて音声、音楽、雑音区間の判別を行うことを特徴とする音声／音楽区間判別手段を具備するオーディオ情報分類装置。

【請求項12】 請求項3ないし請求項11のいずれかに記載のオーディオ情報分類装置において、

入力されたオーディオ情報から有音部のみを抽出する有音抽出手段は、請求項1または2に示された有音判定手段を用いることを特徴とするオーディオ情報分類装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はオーディオ情報の分類装置に関し、特に符号化されていない元のままのオーディオ情報あるいは符号化されたオーディオ情報のいずれから、簡単かつ高速に、音声区間と音楽区間、あるいは音声区間と音楽区間と雑音区間を分類できるオーディオ情報の分類装置に関する。

【0002】

【従来の技術】インターネットに代表されるように、分散したデータベースに、テキストのみならず音声や映像情報が蓄積される技術分野においては、マルチメディア情報を効果的にインデックスする方法が必要とされている。このうちオーディオ信号を分類する手法については、オーディオ信号を音楽や音声区間に分類することで、おおまかなインデックスが可能になる。例えば、E. ScheirerとM. Slaneyの"Construction and evaluation of a robust multifeature speech/music discriminator, Proceedings of IEEE ICASSP, pp.1331-1334, 1997ではオーディオ信号について4Hz成分、フレーム間スペクトル差分、パルス検出の3つの特徴パラメータを利用してBayes 決定法などの識別関数により音声と音楽の判別を行っている。

【0003】図13は前記分類を行う手法の説明図である。オーディオ信号Aは4Hz帯域フィルタ21、周波数変換部22、およびサブバンド分割部23に入力する。4Hz帯域フィルタ21はオーディオ信号Aの4Hz成分を抽出し、4Hz帯域エネルギー演算部24に出力する。周波数変換部22はオーディオ信号Aをスペクトル分析し、フレーム間スペクトル差分演算部25に出力する。また、サブバンド分割部23でサブバンド分割されたオーディオ信号Aは、包絡線ピーク検出部26に出力される。

【0004】一般に、4Hz成分については、音声信号ではこの周波数成分が特に強く出現する特徴がある。フレーム間のスペクトル差分については、音楽のように変化の激しい場合に大きくなる特徴がある。さらに、パルス検出は入力信号を各周波数帯域（サブバンド）に分け包絡線のピークを検出する。音楽のようにリズムのあるオーディオ信号では全ての帯域において周期的にこのピークが現れる。

【0005】オーディオ情報識別部27は、前記の4Hz成分、フレーム間スペクトル差分、パルス検出の3つの特徴パラメータを利用して、Bayes 決定法などの識別関数により音声と音楽の判別を行う。なお、入力してくるオーディオ信号が圧縮符号化されたオーディオ信号である場合には、図示されていない復号処理部で復号して、前記4Hz帯域フィルタ21、周波数変換部22およびサブバンド分割部23に送出する。

【0006】

10 【発明が解決しようとする課題】しかしながら、前記の従来技術は、圧縮符号化されたオーディオ信号から音楽区間、音声区間を検出する場合には、一旦圧縮されたデータを復号してアナログのオーディオ信号Aに戻してから検出処理を行うことになり、処理時間も大幅に増加するという問題点がある。

【0007】また、雑音成分の除去についてはなんら触れておらず、雑音成分を除いた音声や音楽区間の判別ができないという問題がある。換言すれば、雑音成分も音声や音楽区間に含まれてしまうという問題がある。

20 【0008】本発明の目的は、前記した従来技術の問題点に鑑み、符号化されていない元のままのオーディオ情報あるいは圧縮符号化されたオーディオ情報のいずれから、簡単かつ高速に、無音／有音区間の判別、音楽区間と音声区間、あるいは音楽区間と音声区間と雑音区間に分類することを可能とするオーディオ情報分類装置を提供することにある。他の目的は、雑音区間を除去して、音楽区間、音声区間を検出することが可能なオーディオ情報分類装置を提供することにある。

【0009】

30 【課題を解決するための手段】前記した目的を達成するために、本発明は、入力されたオーディオ情報から単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、抽出した単位時間ごとの周波数データのエネルギーの分散を求め、分散値の大きさにより無音／有音区間を判定する無音／有音判定手段とを具備した点に第1の特徴がある。

【0010】また、入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と、有音区間における音の疎密度により音声であるか音楽であるかを判定する音声／音楽区間判定手段とを具備した点に第2の特徴がある。

40 【0011】また、入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と、入力されたオーディオ情報から単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、オーディオ周波数データから単位時間における周波数の重心の平均と重心の標準偏差を求め、周波数の重心の分布により雑音区間か否かを判別する雑音区間抽出手段を具備した点に第3の特徴がある。

50 【0012】さらに、オーディオ周波数データの単位時

5

間における疎密度および単位時間における周波数の重心の平均と重心の標準偏差を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音声と音楽と雑音の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて音声、音楽、雑音区間の判別を行う音声／音楽／雑音区間判別手段を具備した点に第4の特徴がある。

【0013】さらに、オーディオ周波数データの単位時間における周波数の重心の平均と重心の標準偏差を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して雑音と雑音以外の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて雑音と雑音以外の区間の判別を行うことを特徴とする雑音区間判別手段と、オーディオ周波数データの単位時間における疎密度を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音声と音楽の共分散行列を求めておき、前記、雑音区間判別手段で雑音以外と判別された区間に対して、正規分布パターンにおけるBayes 決定識別関数を用いて音声、音楽、雑音区間の判別を行うことを特徴とする音声／音楽区間判別手段を具備した点に第5の特徴がある。

【0014】本発明によれば、符号化されていないものとそのままのオーディオ情報、あるいは符号化されたオーディオ情報のいずれからも、簡単かつ高速に、無音／有音区間の判別、あるいは音声区間、音楽区間、雑音区間を分類することが可能になる。

【0015】

【発明の実施の形態】以下に、図面を参照して、本発明を詳細に説明する。この実施形態は動画像および音声符号化の国際標準であるMPEG1 (ISO/IEC 11172) および MPEG2 (ISO/IEC 13818) により圧縮されたオーディオ符号化データを用いて音声、音楽、雑音区間を分類するものであるが、本発明はこれに限定されるものではない。図1は本発明のオーディオ情報分類装置の一実施形態のブロック図を示す。また、図2は本実施形態の動作を説明するフローチャートである。

【0016】図1に示されているように、圧縮符号化されたオーディオ符号化データaは可変長復号部1に入力される。ここで、圧縮符号化されたオーディオの符号化データ構造について、MPEG1 レイヤーIIを例にして図4を参照して説明する。MPEG1 では図示されているように、元のオーディオ信号pからサンプリングした512個のPCMサンプルをサブバンド符号化して、32個のサブバンドデータ $P_i(n)$ ($n=0, 1, \dots, 31$) を作り、それを時間的にサンプルをずらしながら36回 ($i=0, 1, \dots, 35$) 繰り返して合計1152個のサブバンドデータを作り、この1152個のサブバンドデータを1フレームの符号化データQとしている。

【0017】前記した構造の符号化データQが前記可変長復号部1に連続して入力してくると、該可変長復号部

6

1にはこれを各フレームのサブバンドデータに復号し、サブバンドデータサンプリング部2に出力する。いま、ある単位時間を1秒とすると、この1秒は図5のaのように38フレームから構成されているので、可変長復号部1は1秒分の符号化データに対し、同図のbのように38個の32サブバンド \times 36サンプルを出力する。

【0018】サブバンドデータサンプリング部2では、図5のcに示されているように、単位時間（例えば1秒）分のサブバンドデータのうち、各フレームiのj番目 ($j=0, 1, \dots, 35$ は1フレーム内のサンプル数) にあるサブバンドデータ $S_{ij}(n)$ ($i=0, 1, \dots, 37$ は単位時間内のフレーム数) を抽出し、図1のエネルギー分散演算部3およびメモリ5に入力する。該サブバンドデータサンプリング部2は、入力されたオーディオ情報から単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と呼ぶことができる。

【0019】以上の動作は、図2では、ステップS1～S9で行われる。ステップS1では、フレーム番号を表すiが0と置かれ、ステップS2ではサブバンド番号を表すnが0と置かれる。ステップS3では、可変長復号部1にて符号化データが可変長復号され、ステップS4ではiフレーム目のjサンプル目のサブバンドデータ $S_{i,j}(n)$ が抽出される。次に、ステップS5にて $n=32$ が成立するか否かの判断がなされ、この判断が否定のときはステップS6に進んでnに1が加算される。そしてステップS3に戻って前記と同様の処理が行われる。以上のステップS3～S6の処理が繰り返して行われて、ステップS5の判定が肯定となると、サブバンドデータサンプリング部2から、フレームi、サンプルjのサブバンドデータ $S_{i,j}(n)$ が抽出されたことになる。

【0020】ステップS5の判断が肯定になるとステップS7に進み、iに1が加算される。次にステップ8に進み、 $i=N_f$ が成立するか否かの判断がなされる。ここで、 N_f は単位時間内のフレーム数である。この判断が否定の場合はステップS2に戻り、再び $n=0$ とされて、再度前記した処理が行われる。以上の処理が繰り返し行われ、ステップS8の判断が肯定になると、 $i=0 \sim (N_f - 1)$ フレームの各j番目のサンプルのサブバンドデータ $S_{i,j}(n)$ が抽出されたことになり、ステップS9にてこれらのサブバンドデータ $S_{i,j}(n)$ は図1の各フレームのエネルギー分散演算部3およびメモリ5へ転送される。

【0021】エネルギー分散演算部3では、図6の

(1) および (2) 式に従って、単位時間当たりのエネルギー分散 σ_e^2 を計算し、有音判定部4に入力する。なお、(1) 式で、 N_f は単位時間内のフレーム数、 N_j は1フレーム中のサンプル数で、例えば N_j を1とした場合、フレーム中の先頭のサンプルのみを用いて計算することになり、処理の高速化を図ることが可能である。また、サブサンプルデータ $S_{i,j}(n)$ で $n=0$ と

すると、低周波成分のみを用いてエネルギー分散 σ_e^2 を計算することになり、この場合、高周波成分までを含んだ場合と同等な結果が得られ、処理時間も高速化することが可能である。

【0022】有音判定部4では、入力された単位時間における音声情報が無音であるか有音であるかを下記の(3)式にしたがって判定し、条件に合う場合は有音であると判定する(ステップS11)。有音である場合は、無音である場合に比べて、単位時間のエネルギー分散が大きいから下記の(3)式が成立することになる。

$$\sigma_e^2 > \alpha \quad (3)$$

ここに、 α は予め定められた第1の閾値である。

【0023】該有音判定部4において、入力された単位時間のオーディオ情報が有音であると判断された場合には、メモリ5から該単位時間内の周波数データすなわちサブバンドデータ $S_{i,j}(n)$ を読み出して、全サブバンドエネルギー和演算部7(図3のステップS12)とサブバンドエネルギー重心演算部11(ステップS16)に入力する。この機能は、オーディオ周波数データ抽出手段と呼ぶことができる。一方、無音であると判定された場合には、以降のオーディオ情報判定処理を終了し、ステップS1に戻る。

【0024】全サブバンドエネルギー和演算部7では、図6の(4)式に従って、全サブバンドのエネルギー和 $SE(i,j)$ を計算し、2値化演算部8(ステップ13)に入力する。 $SE(i,j)$ は32バンド分の $S_{i,j}(n)$ のエネルギーの累積和である。2値化演算部8では、図6の(5)式に従って、 $Th1$ を基に $SE(i,j)$ を2値化して、数列 $NS(i,j)$ を計算する。 $Th1$ はあらかじめ定められた2値化のための閾値である。

【0025】音声と音楽の波形は図8のように、音声では断続した波形を持つのに対して、音楽では連続的な波形となる。これらの波形を2値化(正規化)すると、図8の右側の図から明らかなように、音の断続性がより明確になる。すなわち、有音区間における音の疎密度により音声であるか音楽であるかを判定できる。

【0026】2値化演算部8で得られた2値化数列 $NS(i,j)$ は2値化数列分散演算部9(図3のステップS14)に入力する。2値化数列分散演算部9では、2値化数列の単位時間内の分散 σ_s^2 を、図6の(6)式に従って計算し、オーディオ情報識別部10に入力する(ステップS15)。 σ_s^2 は $NS(i,j)$ が0となるサンプル数の分散で、音声区間では断続性が強いいため、該分散値は音楽区間に比べて大きくなる。この分散は、音の疎密度を表している。

【0027】図6の(6)式で、 M は $NS(i,j)$ が単位時間内に1から0に変化する数で、単位時間内の0連続区間の個数を表す。また、 $Nns(k)$ は $NS(i,j)$ が0の場合の連続数で、音楽のようにリズムがある場合は時間的な変化は小さい。

【0028】サブバンドエネルギー重心演算部11(ステップS16)では、図7の(7)式に従って、フレーム i におけるサブバンド重心 $G(i)$ が計算され、エネルギー重心平均、分散演算部12(ステップS17)に入力する。(7)式で、サブバンドの重心はすべてのサブバンド n について、各フレーム内のサンプル j について計算されるが、エネルギー分散 σ_e^2 の場合と同様に、 $Nj=1$ としても重心値に大きな変化がなく、すべてのサンプルについて計算する場合よりも処理時間を削減することが可能である。

【0029】エネルギー重心平均、分散演算部12では、図7の(8)式および(9)式に従って単位時間内の分散 σ_g^2 とエネルギー重心の平均 Eg が計算され、オーディオ情報識別部10(ステップS18)に入力する。図9は単位時間を1秒としたときのサブバンドエネルギー重心の平均と分散の分布例であるが、歓声などの雑音は、音楽や音声などの他の音源と異なって、ある一定の領域 a に集中している。

【0030】オーディオ情報識別部10では、入力された2値化数列分散 σ_s^2 、サブバンドエネルギー重心平均 Eg および分散 σ_g^2 に対して、既知のBayes決定ルールに基づいた正規分布の場合の識別関数(図7の(10)式)を用いて、雑音、音楽、音声の判別が行われる。ここで、クラスは雑音、音楽、音声の3つのクラスに分類する。また、入力ベクトル x は $(\sigma_s^2, Eg, \sigma_g^2)$ の要素で構成される。なお、(10)式における $mk, ck, p(\omega k)$ は、トレーニングデータを用いて、あらかじめ求めておくことができる。判定は、入力ベクトルに対して、最も大きな $f_k(x)$ を与えるクラス k が求める判別クラスとなり、結果を出力する。すなわち、トレーニングにより予め求められた各クラス(雑音、音楽、音声)のデータ $mk, ck, p(\omega k)$ を(10)式に代入し、これに前記(6)(9)(8)式で求められた入力ベクトル $x(\sigma_s^2, Eg, \sigma_g^2)$ を入れて、各クラスの識別値 $f_k(x)$ を求める。そして、該識別値 $f_k(x)$ の一番大きいクラスが雑音であれば雑音、音声であれば音声、音楽であれば音楽と判定する。なお、オーディオ情報識別部10は、K近傍決定、ゆう度検定、K-平均法、K-決定木法などのような前記(10)式以外の他の式を用いてクラスの判別をするようにしても良い。

【0031】次に、本発明の第2の実施形態について、図10を参照して説明する。図10において、図1と同一または同等物には同じ符号が付されている。図10の可変長復号部1～有音判定部4の動作(図2のステップS1～S11)は前記第1実施形態と同じであるので、説明を省略し、サブバンドエネルギー重心演算部11以降の動作を、図11を参照して説明する。

【0032】有音判定部4において、入力された単位時間のオーディオ情報が有音であると判断された場合に

は、メモリ5から単位時間内のサブバンドデータ $S_{i,j}(n)$ を読み出してサブバンドエネルギー重心演算部11に入力する。一方、無音であると判定された場合には、以降のオーディオ情報判定処理を終了し、ステップS1に戻る。

【0033】サブバンドエネルギー重心演算部11（ステップS16）では、図7の（7）式に従って、フレーム i におけるサブバンド重心 $G(i)$ が計算され、エネルギー重心平均、分散演算部12（ステップS17）に入力する。（7）式で、サブバンドの重心は全てのサブバンド n について、各フレーム内のサンプル j について計算されるが、エネルギー分散 σ_e^2 の場合と同様に、 $N_j = 1$ としても重心値に大きく変化がなく、すべてのサンプルについて計算する場合よりも処理時間を削減することが可能である。

【0034】エネルギー重心平均、分散演算部12では（8）式および（9）式に従って単位時間内の分散 σ_g^2 とエネルギー重心の平均 E_g が計算され、雑音識別部13（ステップS18）に入力する。

【0035】雑音識別部13では、入力されたサブバンドエネルギー重心平均 E_g および分散 σ_g^2 に対してBayes決定ルールに基づいた正規分布の場合の識別関数（10）式を用いて、雑音が否かの判別が行われる。ここで、クラスは雑音と雑音外の2つに分類する。また、入力ベクトル x は (E_g, σ_g^2) の要素で構成される。（10）式における $m_k, c_k, p(\omega_k)$ は、トレーニングデータを用いて予め求めておくことができる。判定は、入力ベクトルに対して、最も大きな $f_k(x)$ を与えるクラス k が求める判別クラスとなり、結果を出力する。

【0036】ここで、雑音と判定された場合（ステップS30が肯定）は、雑音である旨の結果を出力後、最終データでない限り（ステップS23が否定）、次のデータ入力を行う。また、雑音外と判定された場合（ステップS30が否定）は、次の処理（ステップ12）へ進み、音楽か音声の判定を行う。

【0037】音楽か音声の判定処理に進むと、メモリ5から全サブバンドエネルギー和演算部7に $S_{i,j}(n)$ が入力され、全サブバンドエネルギー和演算部7では、図6の（4）式に従って、全サブバンドのエネルギー和 $SE(i,j)$ を計算し、2値化演算部8（ステップ13）に入力する。 $SE(i,j)$ は32バンド分の $S_{i,j}(n)$ のエネルギーの累積和である。2値化演算部8では、図6の（5）式に従って、 $SE(i,j)$ を2値化して、数列 $NS(i,j)$ を計算する。 $Th1$ は予め定められた2値化のための閾値である。

【0038】2値化演算部8で得られた2値化数列 $NS(i,j)$ は2値化数列分散演算部9（ステップ14）に入力する。2値化数列分散演算部9では、2値化数列の単位時間内の分散 σ_s^2 を図6の（6）式にしたがっ

て計算し、音楽音声識別部14に入力する（ステップ15）。 σ_s^2 は $NS(i,j)$ が0となるサンプル数の分散で、音声区間では断続性が強いので、該分散値は音楽区間に比べて大きくなる。

【0039】音楽音声識別部14では、入力された2値化数列分散 σ_s^2 に対してBayes決定ルールに基づいた正規分布の場合の識別関数（10）式を用いて、音楽、音声の判別が行われる。ここで、クラスは音楽、音声の2つのクラスに分類する。また、入力ベクトル x は (σ_s^2) の要素で構成される。さらに、（10）式における $m_k, c_k, p(\omega_k)$ は、トレーニングデータを用いて、予め求めておくことができる。判定は、入力ベクトルに対して、最も大きな $f_k(x)$ を与えるクラス k が求める判別クラスとなり、結果を出力する。

【0040】以上のように、前記第1、第2実施形態によれば、圧縮符号化されたオーディオの符号化データから無音/有音を判別し、有音の場合、音楽区間、音声区間、雑音区間を区別し、それぞれのタイムコードを図示されていない音声区間保持部、音楽区間保持部、雑音区間保持部のそれぞれに記録させることができる。

【0041】さらに、本発明は圧縮されていないオーディオ情報の分類に関しても適用できる。その場合の実施形態を以下に説明する。

【0042】圧縮符号化されていないオーディオ情報を扱う場合は、図1の可変長復号部1およびサブバンドデータサブサンプリング部2は高速フーリエ変換部（以下FFT変換部）に置き換えられる。元のオーディオ情報からこのFFT変換部において、図12にあるようなFFT変換を行い、単位時間分の周波数データを抽出する。今、該単位時間を1秒とすると、元のオーディオ信号 p からサンプリングした2048個のサンプルをFFT変換し、それを時間的にサンプルをずらしながら38回繰り返して合計2048×38個のFFTデータを単位時間分の周波数データとしている。

【0043】その後、各フレームのエネルギー分散、エネルギー重心演算の平均および分散、エネルギー和の2値化後の数列分散を計算して、無音/有音、音楽、音声、雑音の判定を行う。

【0044】

【発明の効果】以上の説明から明らかなように、本発明によれば、圧縮符号化されたあるいは圧縮符号化されていないオーディオデータから、符号化データ上で、オーディオ情報を有音/無音、音楽/音声/雑音区間に分類することが可能である。

【0045】本発明を実際に動作させ、MPEG1レイヤIIで符号化された15分間のテレビ番組を用いて1秒毎の分類を行ったところ、無音の判定は92%、音声区間の検出は99%、音楽区間は75%、雑音区間は74%程度検出することが可能になった。

【図面の簡単な説明】

【図1】 本発明の一実施形態の構成を示すブロック図である。

【図2】 本実施形態の動作を示すフローチャートである。

【図3】 図2の続きのフローチャートである。

【図4】 MPEGオーディオ符号化データの構造を説明するための図である。

【図5】 図1のサブバンドデータサブサンプリング部の動作を説明するための図である。

【図6】 本実施形態で使用される数式を表す図である。

【図7】 本実施形態で使用される数式を表す図である。

【図8】 音声および音楽の正規化前および正規化後の波形図である。

【図9】 雑音のサブバンド重心の平均を表す図である。

【図10】 本発明の第2実施形態の構成を示すブロック図である。

【図11】 第2実施形態の要部の動作を示すフローチャートである。

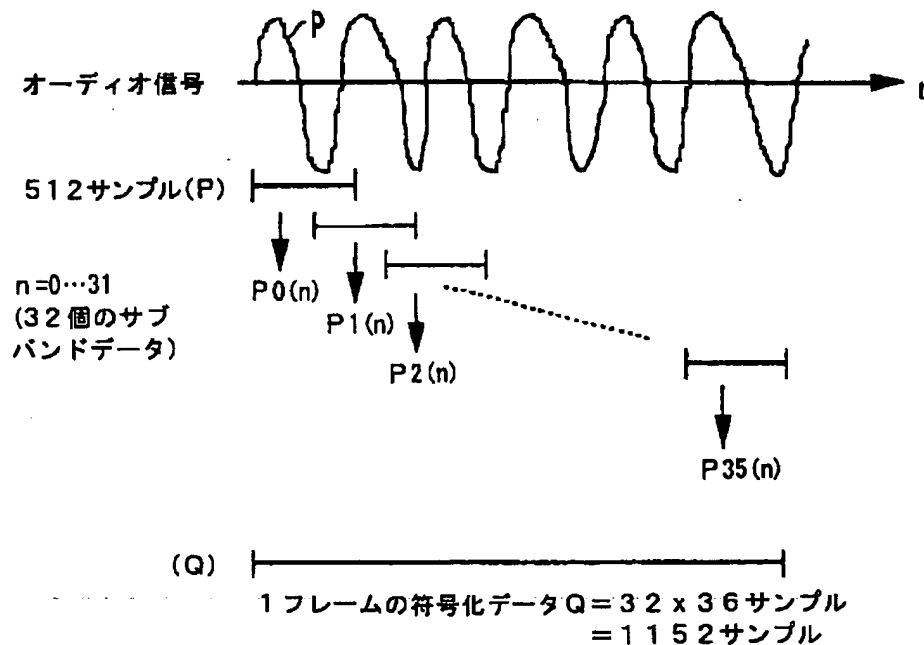
【図12】 符号化されていないオーディオ情報の周波数データの抽出方法を説明するための図である。

【図13】 従来のオーディオ情報分類装置の構成を示すブロック図である。

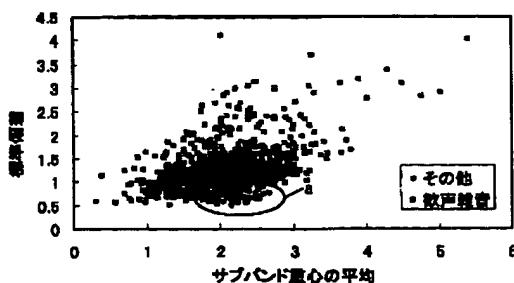
【符号の説明】

10 1…可変長復号部、2…サブバンドデータサブサンプリング部、3…エネルギー分散演算部、4…有音判定部、5…メモリ、6…制御部、7…全サブバンドエネルギー和演算部、8…2値化演算部、9…2値化数列分散演算部、10…オーディオ情報識別部、11…サブバンドエネルギー重心演算部、12…エネルギー重心平均・分散演算部、13…雑音識別部、14…音楽音声識別部。

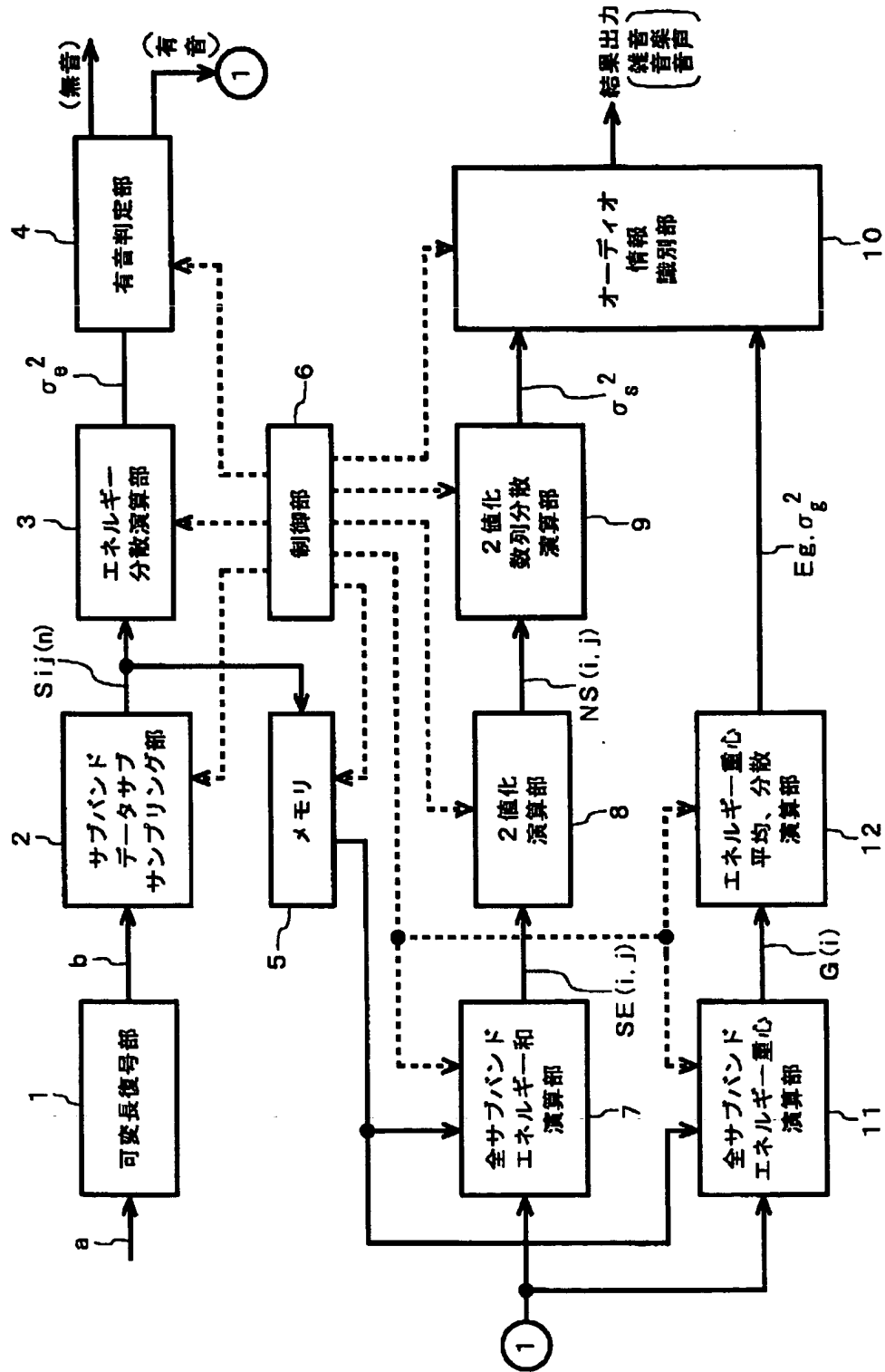
【図4】



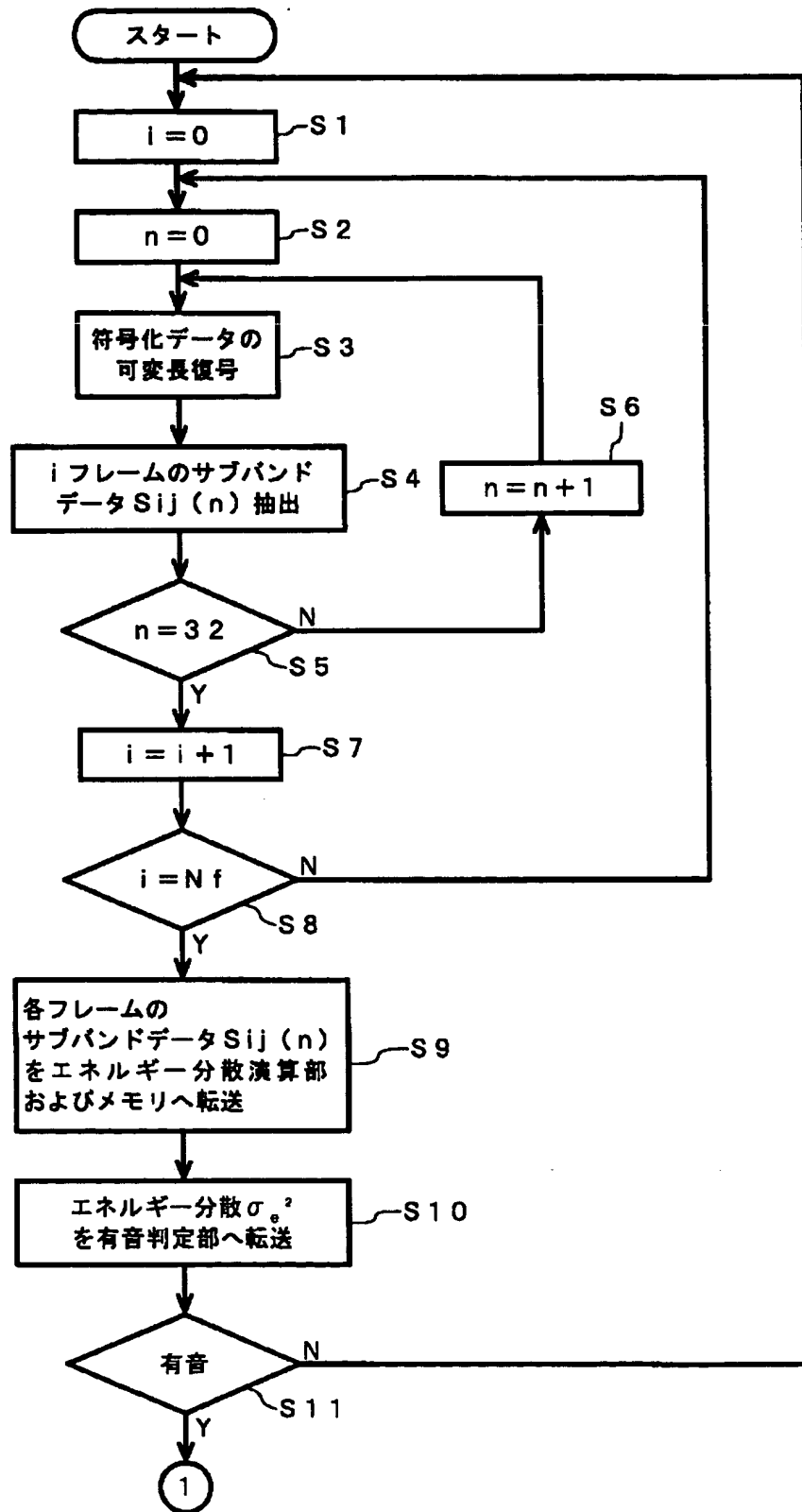
【図9】



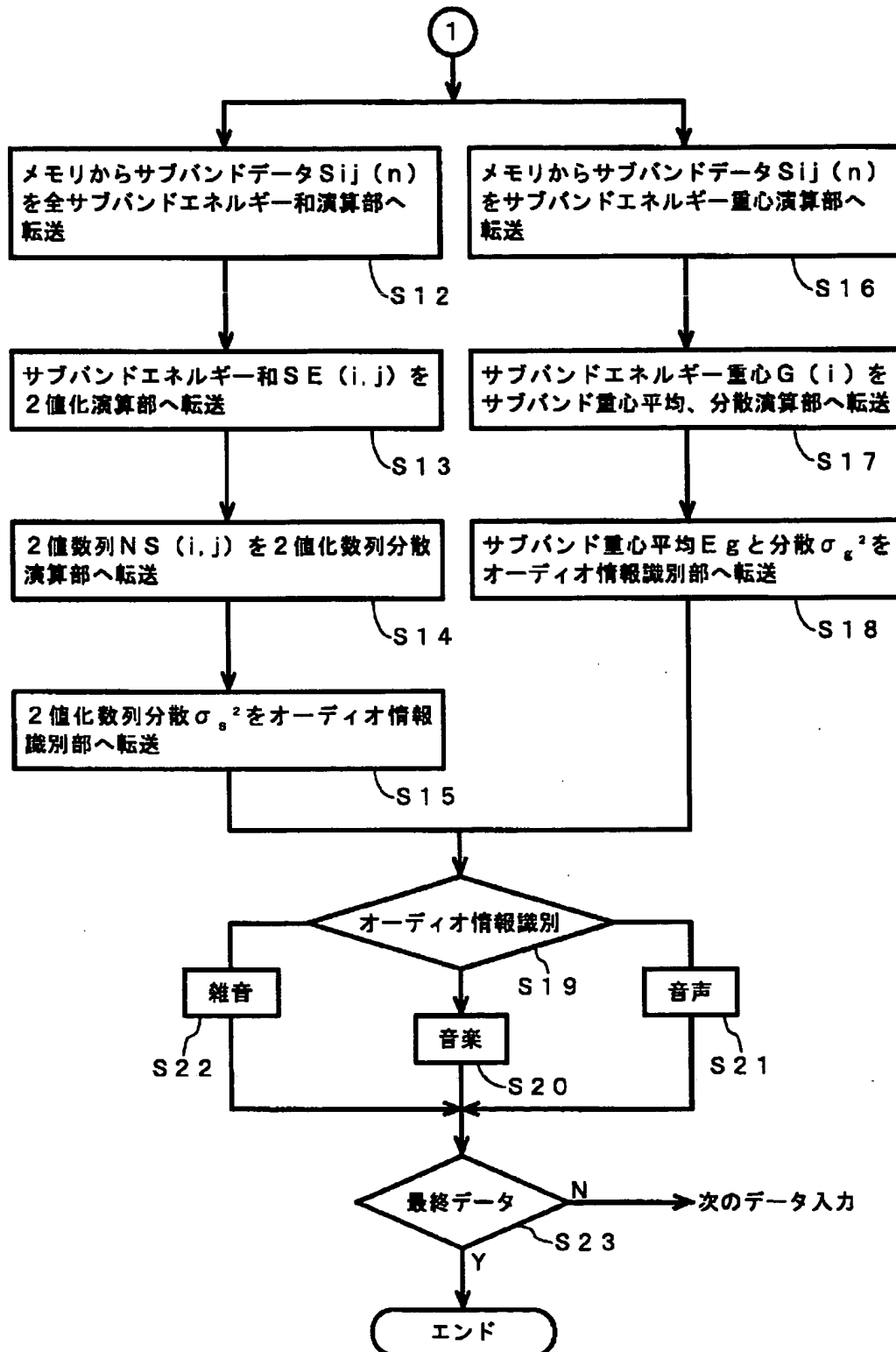
【図1】



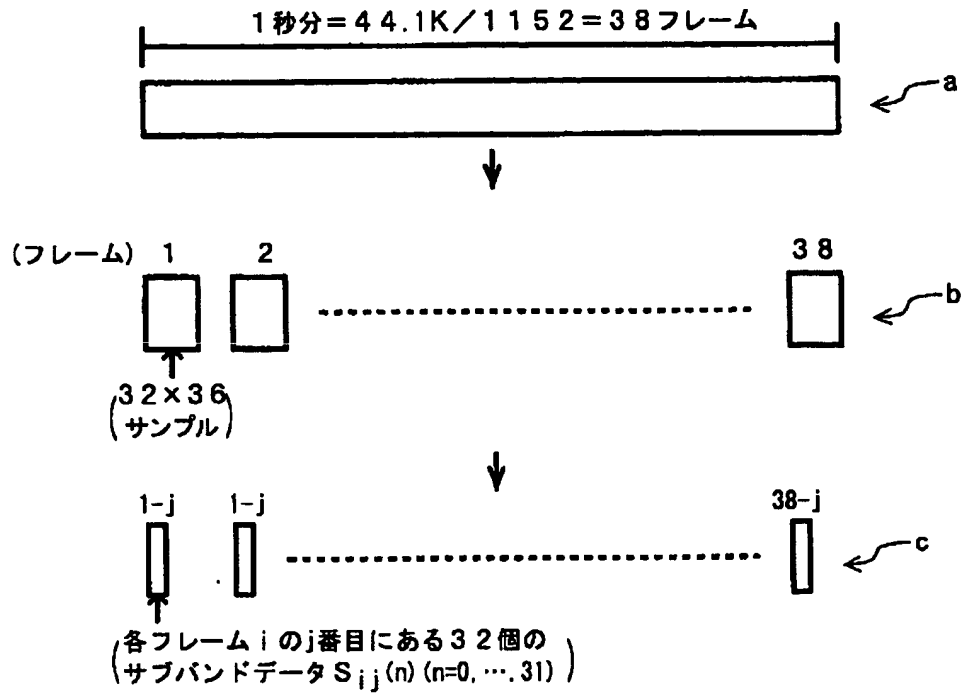
【図2】



【図3】



【図5】



【図7】

$$G(i) = \frac{\sum_{n=0}^{31} \sum_{j=0}^{Nj-1} |S_{i,j}|^2(n)}{\sum_{n=0}^{31} \sum_{j=0}^{Nj-1} S_{i,j}^2(n)} \quad (7)$$

$$\sigma_g^2 = \frac{1}{Nf} \sum_{i=0}^{Nf-1} (G(i) - Eg)^2 \quad (8)$$

$$Eg = \frac{1}{Nf} \sum_{i=0}^{Nf-1} G(i) \quad (9)$$

$$f_k(\vec{x}) = -\frac{1}{2} (\vec{x} - \vec{m}_k)^T C_k^{-1} (\vec{x} - \vec{m}_k) + [\log p(w_k) - \frac{1}{2} \log |C_k|] \quad (10)$$

\vec{x} : 入力ベクトル ($\sigma_s^2, Eg, \sigma_g^2$)
 \vec{m}_k : クラスkの平均値ベクトル
 C_k : クラスkの共分散行列
 $p(w_k)$: クラスkの発生確率

【図6】

$$\sigma_{\theta}^2 = \frac{1}{NfNj} \sum_{i=0}^{Nf-1} \sum_{j=0}^{Nj-1} ((S_{i,j}^2(n) - \langle S_{i,j}^2(n) \rangle)^2) \quad (1)$$

$$\langle S_{i,j}^2(n) \rangle = \frac{1}{NfNj} \sum_{i=0}^{Nf-1} \sum_{j=0}^{Nj-1} S_{i,j}^2(n) \quad (2)$$

$(S_{i,j}(n)$ はフレーム*i*, サンプル*j*, サブバンド*n*のサブバンドデータ
 Nf は単位時間あたりのフレーム数
 Nj は1フレーム中のサンプル数

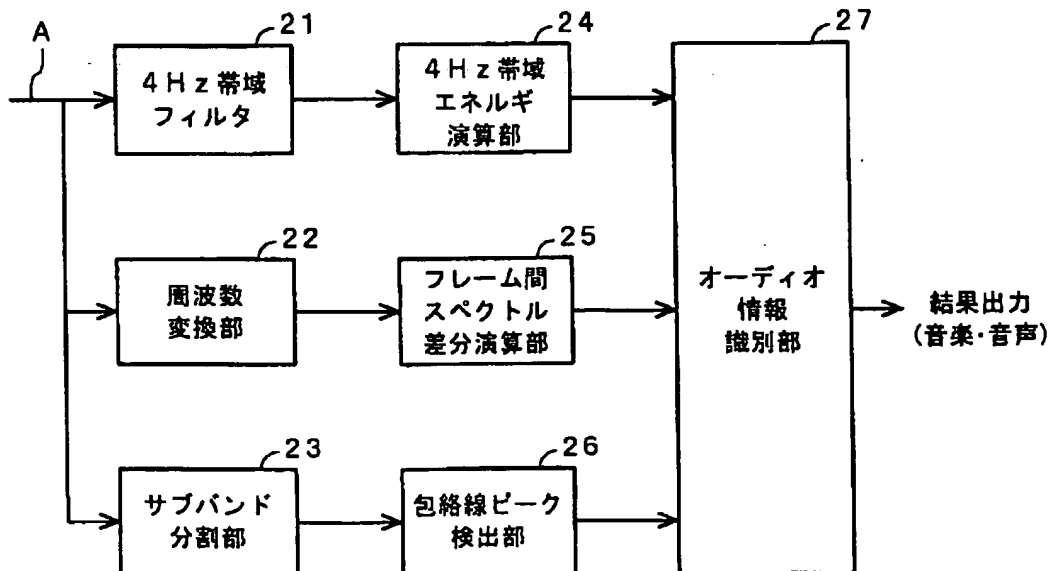
$$SE(i,j) = \sum_{n=0}^{31} S_{i,j}^2(n) \quad (4)$$

$$NS(i,j) = \begin{cases} 1 & \text{if } SE(i,j) > Th1 \\ 0 & \text{else} \end{cases} \quad (5)$$

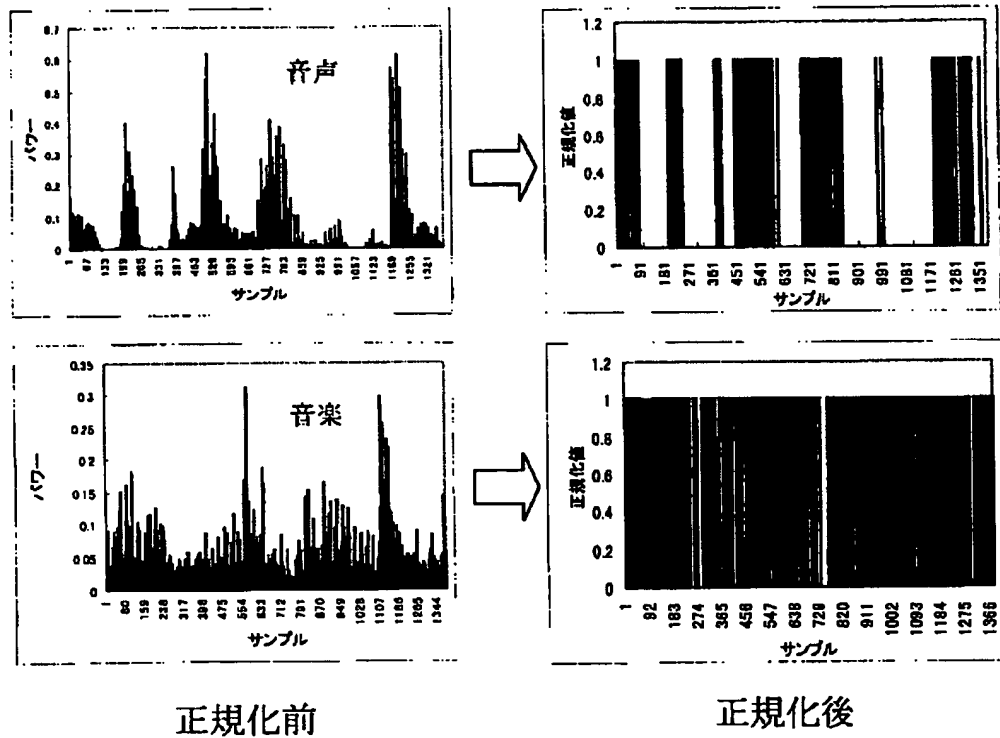
$$\sigma_s^2 = \frac{1}{M} \sum_{k=0}^M (N_{ns}(k) - \langle N_{ns}(k) \rangle)^2 \quad (6)$$

$(M$ は $NS(i,j)$ が単位時間内に1から0へ変化する数
 $N_{ns}(k)$ は $NS(i,j)$ が0となる連続数
 $\langle N_{ns}(k) \rangle = \frac{1}{M} \sum_{k=0}^M N_{ns}(k)$

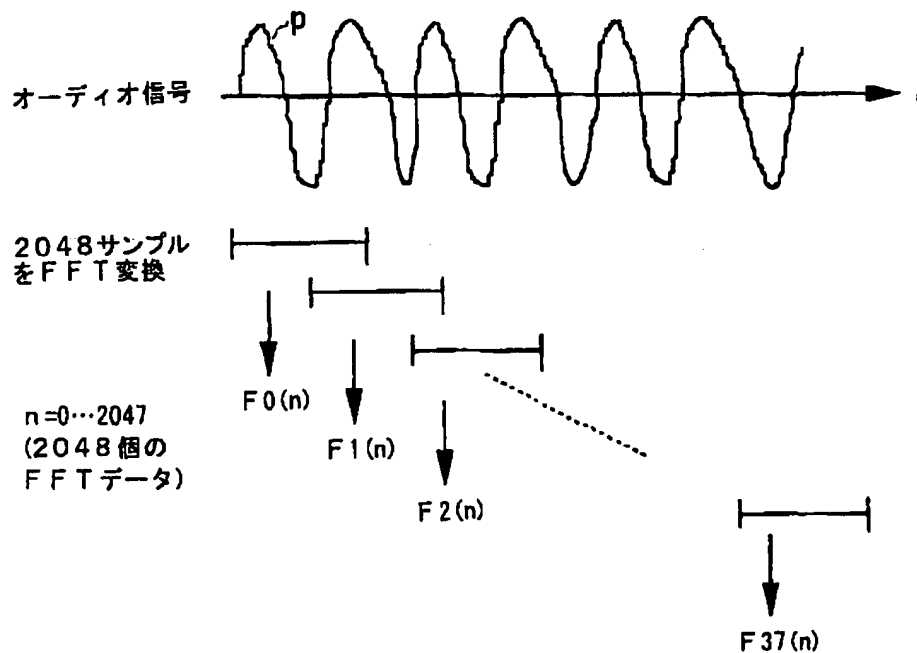
【図13】



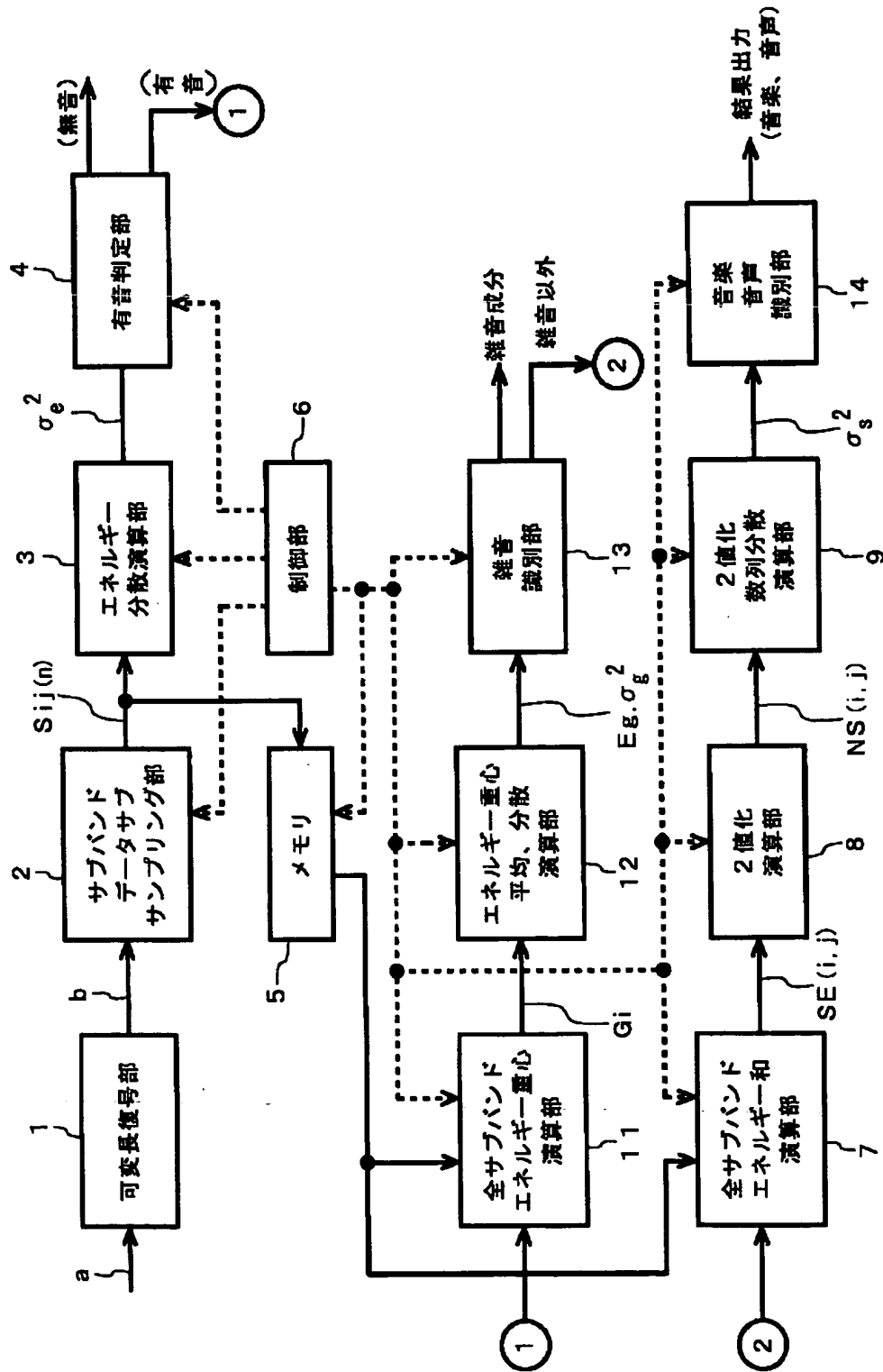
【図8】



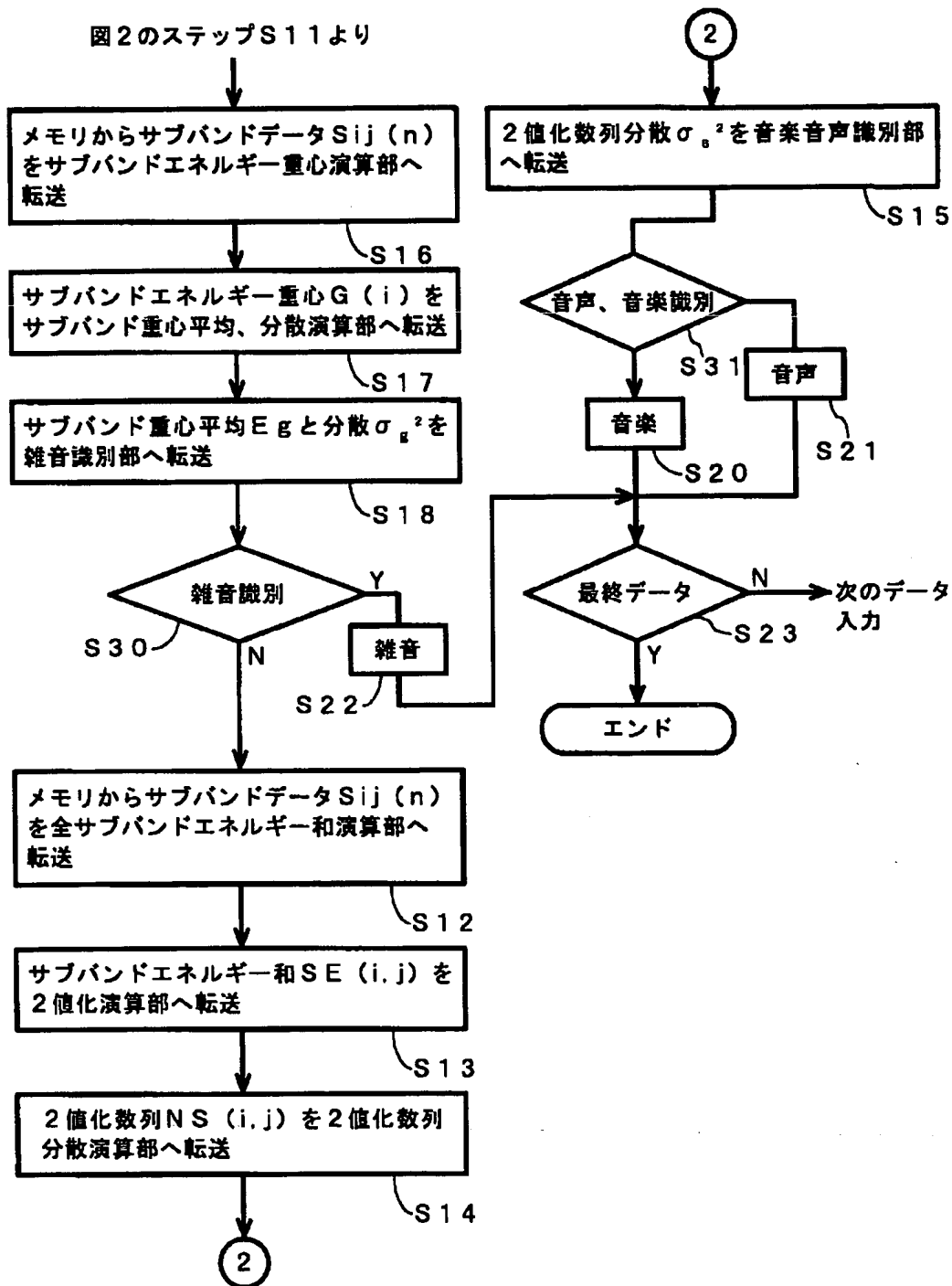
【図12】



【図10】



【図11】



フロントページの続き

(72)発明者 米山 暁夫
東京都新宿区西新宿2丁目3番2号 国際
電信電話株式会社内

(72)発明者 柳原 広昌
東京都新宿区西新宿2丁目3番2号 国際
電信電話株式会社内